

Plaintiffs' Exhibit 138

First look detection at the rule and iu level

Date: October 8, 2015

Authors: Jim Giles <jjimgiles@google.com>, tbd...

DRAFT IDEA

First look detection at the rule and iu level

Goal

Background

High Level Designs

IU/rule Detection

Limitations of IU/rule-level Detection

How this would work

Other considerations

Page-level Detection

Limitations of iu/rule and page-level detection techniques

NOTE -- This is a DRAFT/WIP -- there are details to work on IU/rule mapping and analysis to be done to see if it is even worth doing this way

Goal

We would like buyers on the Ad Exchange to understand if they are bidding on "First Look" inventory or if they are bidding on inventory that may have been seen by other buyers previously. "First Look" inventory is more valuable to buyers, and hence should result in higher bids while inventory that may have been seen by other buyers is less valuable and should result in lower bids.

Background

We have several parallel efforts to detect or declare first look at the query level (see for example [go/mediation-review](#) and [go/gpt-firstlook](#)). These efforts each have made great progress, but have limitations. For example, our online mediation detection approaches are currently able to precisely identify mediation less than 10% of the time. We have promising improvements for mediation detection (eg, using Mawler), but there will always be limitations. The Javascript declaration described in [go/gpt-firstlook](#) is also promising because publishers have control of the tags and can be explicit about the first look and mediation state. However, this is subject to human error and fraud. More importantly, it requires retagging so adoption will be slow. Publishers will be incented to declare first look, but will have a strong disincentive to declare mediation.

High Level Designs

IU/rule Detection

Rather than attempting to declare first look on a query by query basis, we will declare "first look starts" and "mediation starts" on a inventory unit / rule. We will declare an iu as "First Look Starts" if and only if *none* of its associated active line items are mediated. We will declare an iu as "Mediated Starts" if *any* of its associated active line items are mediated. A rule will be "First Look" if it targets only IUs that are first look and it will be "Mediated starts" if any IU it targets is mediated. We will use the same iu/rule declaration for *all* queries regardless of whether a given query is first look or comes as a passback from a mediator until there is a configuration change. This means that if a publisher wants a "First Look" declaration for an iu/rule and the associated higher expected real-time bids, the publisher should not set up any line items with mediators.

We would automatically determine the First Look Starts / Mediation Starts status based on the configuration. We would create a global opt-in setting for each publisher, or surface it as an opportunity. We might want to allow opt in at the creative level as well, but this is not necessary.

Limitations of IU/rule-level Detection

For this solution, "First Look Starts" and "Mediated Starts" are associated with iu/rules rather than queries. As such, we would miss an opportunity to offer more fine-grained information and reward individual first look queries -- it is all or nothing for every query associated with the iu/rule.

For large publishers who have mediators configured for most of their iu/rules, we need to tread carefully. On the one hand, we could try to convince publishers to remove mediators from slots to get the First Look Starts designation and associated higher bids. On the other hand, if they keep mediators configured for an iu/rule, we would mark all such queries as Mediated Starts, even for queries that come to Ad Exchange first. We would offer lower bids for all of these queries, making Ad Exchange look less attractive than other exchanges. This might still be an incentive to remove mediators though.

This solution doesn't help for the case where the mediator has the tag on the page, because the First Look Starts or Mediated Starts designation is derived from the DFP configuration rather than the query flows. In other words, this technique can't tell us about what happens before the query reaches DFP. We could, however, combine anything determined to be Mediated Starts through our other techniques with this signal to capture more of these cases.

How this would work

Assuming the publisher has allowed it (global opt in, opportunity, or per ad-slot opt in), we would classify each iu/rule, geography pair as follows:

For each iu/rule:
 For each associated line item:
 if snippet contains known mediator:

Comment [1]: Could you please make it explicit that we are focusing here on the tag on the page that is the first tag in the call sequence, that is to say mediation = STARTS a call that results in a mediation chain. I think it'll help readability

Comment [2]: sure!

Comment [3]: A refinement may be needed for the (common ?) case where the first call comes to XFP and the second to AdX. We need to figure out the AdX adslot code used in the mediated query. Is it available somewhere in the account settings.

Comment [4]: The pub will have an AdX account. One issue I foresee is that they might be using the same 'dummy' adslot code for all XFP IUs.

Comment [5]: Yeah, this is something we have to check -- I'm assuming that there is a way to map from an adslot/url in dfp to an adslot/url in adx. In that case, we'd keep the complementary map in adx. If not, then it might be back to the drawing board.

Comment [6]: this map should be buildable by looking at offline crawls, right?

Comment [7]: Yeah, good point!

Comment [8]: To extract this elasing the crawl has to first establish that mediation is happening. So we'll already have the information we're seeking, i.e., whether the adslot is ever mediated.

Comment [9]: What we're seeking is a more robust way to extract the information directly from account settings.

Comment [10]: One possibility could be to use a hybrid approach - combine the adslot mediation data from account settings and Mawler crawls. Use information from wherever we can get it. But first we need to understand what kind of information is available from each source - both in terms of precision (can we construct a decent IU->adslot map) and coverage (Mawler)

Comment [11]: is there already a way/plan as to how to do that? The creative could be basically anything, including arbitrary code. Are we talking about a lookup for mediator names in the code?

Comment [12]: That's what I was thinking of - Zach said they can already fully classify 60% of entities -- he thought finding a mediator in the snippet would be possible at a much higher precision

Comment [13]: I see, but that still relies on the whitelist then (and on a manual lookup of the used domain names)

set (iu/rule/ line item geography) = Mediated Starts
If no associated line item snippet has a mediator domain for a given geography:
set (iu/rule / geography) = First Look Starts

Geography would be limited to world/country initially, but could be finer grained.

We would make these iu/rule / geography / status maps available at serving time. As queries arrive, we will use the iu/rule and derived geography from the query to look up the mediation status in the iu/rule / geography map. The status would be present in the backfill call, and ultimately sent out to RTB buyers (repurposing our current mediation callout field).

Other considerations

We'd probably want to monitor the snippets for new domain names that might be mediators.

Page-level Detection

For XFP-AdX style mediation the iu/rule level detection methodology formulated above relies on offline Mawler scans to connect XFP IUs sent as part of the mediated request to AdX. However involving detection-based methods is contrary to the philosophy behind mediation self-declaration and we would like to avoid it as far as possible.

An alternative to iu/rule level detection is page-level detection wherein we define an AdX query as mediated if

- the query originates from a URL associated with a line item having a mediator tag
- the query is non-backfill
- there does not exist an AdX tag on the page
 - there is a corner case where pubs do backfill via direct AdX calls. If that volume is too high we may need to involve Mawler

Limitations of iu/rule and page-level detection techniques

- Involve Mawler, i.e., some aspects of mediation detection
- May generate too many false positives defeating the purpose of self-declaration. As an example, consider the following scenario
 - a. Pub wants to calibrate the effect of mediation via self-declaration
 - b. To turn off mediation the pub lowers the priority of a line item with a mediator tag so that it does not get first look
 - c. Our analysis does not take priorities into account and will still pick up mediation because it still sees a line item with mediators.
 - Can we take priorities into account ? Probably, but then the approach suddenly gets more complex. What all is there to consider ?
- **Bottomline:** these techniques do not provide an adequate level of control to the pub to govern mediation behavior

